Hyperspectral Image Classification Based on Three-Dimensional Dilated Convolution and Graph Convolution

[1st] Huanhuan Lv, [2nd] Shuang Bai, [3rd] Hui Zhang *
lvhh2010@126.com, 1023650988@qq.com, * Corresponding author: 03013@zjhu.edu.cn

[1]College of Software Liaoning Technical University Huludao, Liaoning, China
College of Information Engineering Huzhou University Huzhou, Zhejiang, China

[2]College of Software Liaoning Technical University Huludao, Liaoning, China
College of Information Engineering Huzhou University

[3]College of Information Engineering Huzhou University Huzhou, Zhejiang, China

Abstract—Hyperspectral images with high dimensionality, strong inter-band correlation and high spectral resolution make the research of existing classification methods a great challenge. Typical convolutional neural network models cannot capture the feature information of irregular or inhomogeneous regions of feature classes in images, and the features extracted by using a single network model lack diversity and cannot provide the best classification results. To address the above problems, a hyperspectral image classification method that fuses 3D dilated convolution and graph convolution is proposed. Firstly, a 3D dilated convolution network model is constructed to extract multi-scale null-spectral features using dilated convolution with different dilated parameter sizes; secondly, a neighborhood relationship-based graph convolution neural network model is established to obtain spatial structure contextual features by aggregating the neighborhood feature information of graph nodes; then, to improve the feature representation capability, the extracted deep null-spectral features are fused with spatial contextual features Finally, the proposed method is compared and analyzed with seven related classification methods on two hyperspectral datasets, Indian Pines and Pavia University. The results show that the overall classification accuracy, average classification accuracy and kappa coefficient of the proposed method reach 99.27%, 98.81%, 99.12% and 99.32%, 99.12%, 98.38%, respectively, on the two datasets, indicating that the proposed method can make full use of the diverse features of hyperspectral images and has a strong feature learning capability, which effectively improves the classification accuracy of images.

Keywords—hyperspectral image classification; three-dimensional convolutional neural network; dilated convolution; graph convolution network; feature fusion

ITEI 2022, December 23-25, Harbin, People's Republic of China

DOI 10.4108/eai.23-12-2022.2329104

1 INTRODUCTION

Hyperspectral remote sensing is a technology that can make full use of visible, infrared and microwave light for high-altitude detection and remote non-contact classification and identification of land features through information processing and transmission. The acquired hyperspectral images have hundreds of adjacent narrow bands containing a large number of channel dimensions, which can provide rich spectral and spatial information to effectively improve the identification of land cover types[1]. Therefore, hyperspectral images have been widely used in various fields such as agriculture[2] and environmental monitoring[3]. Hyperspectral image classification, which refers to the use of its unique spatial spectral features to assign a unique class label to each image element in an image, is an important research aspect in many application areas. However, the problems of high spectral dimensionality, high inter-band correlation, large data redundancy, and small amount of labeled sample data in hyperspectral images make it a great challenge to achieve high accuracy classification[4,5].

In the early days, methods such as K-nearest neighbor (KNN)[6], support vector machine (SVM)[7], and random forest (RF)[8] were proposed to achieve hyperspectral image classification, but these methods only considered the spectral information and ignored the rich spatial information that hyperspectral images have. The phenomenon of "same-spectrum and different- spectrum" in hyperspectral images[9,10] has led to the inability to achieve accurate classification by using spectral features alone. In hyperspectral images, spatially adjacent image elements have similar spectral features and similar category information with high probability, so adding spatial features to the feature extraction and classification process can effectively reduce the classification uncertainty and improve the classification accuracy. Therefore, methods such as extended morphological profile (EMP)[11], Markov random field (MRF)[12], and Gabor filtering[13]have been used to extract spatial features of hyperspectral images thus improving the classification accuracy. Although the above methods have proven to be effective in improving the classification accuracy of hyperspectral images, only single, shallow spectral features or spatial features are extracted for hyperspectral image classification, which cannot fully realize the feature representation and learning, resulting in less than ideal classification accuracy.

In recent years, deep learning-based classification methods have received attention due to the powerful ability of Convolutional Neural Network (CNN)[14] to automatically extract deep-level features. Several researchers have designed different CNN structures to improve the classification accuracy of hyperspectral images. liu et al[15] proposed a three-dimensional convolutional neural network (3D-CNN) that takes the original hyperspectral image as input to directly extract the null spectral features, which improves the classification accuracy while reducing the number of parameters. roy et al[16] proposed a hybrid 3D-2D CNN (HybridSN) for hyperspectral image classification, which combines the advantages of 3D-CNN and 2D-CNN to extract the null spectral features and achieves good classification results. To better learn the deep null-spectrum features, reduce the network complexity, and avoid the overfitting problem caused by the increase of layers, Jeyaraj et al[17] went to build a deep network to extract the null-spectrum features by residual network (ResNet), which further improved the classification performance of images. Chen Wenhao et al[18] proposed a 3D residual convolutional neural network (SE-ResNet) with the introduction of channel attention mechanism to extract null-spectrum features based on a network with residual structure as the backbone, and obtained

better classification results. In order to avoid the problem of feature information loss caused by pooling layer, Yan Mingjing et al[19] proposed a three-dimensional null convolutional residual neural network (Dilated-3D-CNN) hyperspectral image classification method by introducing null convolution, which used null convolution instead of pooling layer to improve the accuracy of hyperspectral image feature classification without increasing the network parameters. The research of the above methods has some practical significance for the improvement of classification accuracy. However, in the CNN-based model, the convolution or pooling operation relies heavily on the fixed structure in the regular grid, which cannot capture the subtle changes of continuous shapes associated with pixels in the spectral features and ignores the irregular spatial topological relationships of hyperspectral images, resulting in incomplete feature information extraction and easy misclassification.

To be able to effectively extract information from inhomogeneous and irregular regions in images, graph convolutional neural network[20] (GCN) is applied to image processing as a network model with graph structure as input, which can convolve arbitrary irregular image regions flexibly, so it is getting more and more attention in hyperspectral image classification. Wan et al[21] proposed a multi-scale dynamic graph convolution hyperspectral image classification method (MDGCN), which expresses the spatial topological relationships by building multiple feature maps at different neighborhood scales. Xibing Zuo et al[22] proposed a graph convolution network-based hyperspectral image classification method (RULBP + GCN), by constructing graphs based on texture feature extraction and inputting them into the graph convolution network to improve the classification accuracy of images in the case of small samples. Liu et al[23] proposed an enhanced graph convolution network-based hyperspectral image classification method (CEGCN), by using encoder and decoder to fuse CNN with GCN, eliminating the network structure incompatibility problem to improve the classification accuracy. Although the above method is able to extract the spatial topological relations, it does not reasonably utilize the deep spatial spectral features with the contextual features extracted from the spatial topological relations, which leads to feature loss making the classification accuracy cannot be further improved.

To solve the above problems and express the feature information of hyperspectral images completely, this paper proposes a hyperspectral image classification method (Dilated-3D-GCN) based on three-dimensional dilated convolution and graph convolution. The contributions of this method are as follows: the proposed combination of 3D-CNN and dilated convolution (DC) constitutes a 3D-DC module to extract deep null-spectral features of hyperspectral images, which expands the perceptual field while reducing the generation of parameters and provides favorable conditions for the extraction of deep null-spectral features; the topological features of spatial structure in the image are fully extracted by GCN, and graph is performed on the global features obtained by using KNN The deep null-spectral features extracted by 3D-DC and the contextual features extracted by GCN are fused in two channels to achieve complete feature representation, and better performance and higher classification accuracy are obtained in hyperspectral images.

2 METHODOLOGY OF THIS ARTICLE

2.1 Multi-scale feature extraction based on 3D dilated convolution

Hyperspectral image as a kind of 3D data with spectral dimension and spatial dimension, using 3D-CNN can fully extract the spatial information and spectral information. It operates the adjacent bands of the input layer in spatial dimension and spectral dimension simultaneously through 3D convolution kernel to extract the null spectral features. the equation of 3D-CNN convolution calculation is

$$\boldsymbol{V}_{i,j}^{xyz} = f\left(\sum_{m}\sum_{l=0}^{L_i-1}\sum_{w=0}^{W_i-1}\sum_{h=0}^{H_i-1} W_{i,j,m}^{lwh} \bullet \boldsymbol{V}_{i-1,m}^{(x+l)(y+w)(z+h)} + b_{i,j}\right) \tag{1}$$

where m denotes the feature map connected to the current feature map in layer $i-1$; L_i and W_i denote the length and width of the convolution kernel; H_i denotes the size of the convolution kernel in the spectral dimension; W represents the connection weight of the mth feature map connected to layer $i-1$; $b_{i,j}$ denotes the bias of the Jth feature map in layer i; and f is the activation function.

In general, the smaller the convolution kernel is, the smaller the perceptual field is, and the less global information is received; on the contrary, the increase of convolution kernel will increase the perceptual field to obtain more effective information, but the increase of convolution kernel will increase the complexity of the network. To address this problem, this paper introduces the dilated convolution(DC), which expands the perceptual field to make the edge information more complete and the spatial information more adequate while keeping the computation volume of the network unchanged. The perceptual field is calculated as follows:

$$F_i = F_{i-1} + \left(k_i - 1\right) \times d_i \times \prod_{n=1}^{i-1} s_n \tag{2}$$

where F_i is the perceptual field of the convolutional kernel in the ith convolutional layer; k_i denotes the size of the convolutional kernel; d_i is the dilated coefficient; F_{i-1} is the perceptual field of the convolutional kernel in the i-1th convolutional layer; and s_n is the convolutional step size.

In this paper, a 3D-DC network composed of DC is introduced on the basis of 3D-CNN, as shown in Figure 1. The network consists of four parts: the first part contains two convolutional layers with a convolutional kernel size of 3×3×3 and a channel number of 16, one of which is a null convolutional layer; the second part contains three convolutional layers with a convolutional kernel size of 3×3×3 and a channel number of 32, the middle of which is a null convolutional layer; the third part contains three convolutional layers with a convolutional kernel size of 3×3×3 and a channel number of 48, of which two are empty convolutional layers; the fourth part contains two convolutional layers with convolutional kernel size of 3×3×3 and

channel number of 64, and one is a Maxpooling layer. All the convolutional layers have a convolutional step of 1, and the ReLU function is used after batch normalization (BN) in each layer to enhance the generalization ability of the model, while the Dropout regularization method is used to avoid overfitting before classification. Where Conv means convolution layer and DC means dilated convolution.

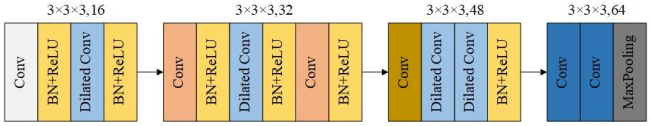

Fig. 1. 3D-DC network structure

2.2 Extraction of Spatial Topological Relations Based on GCN

There are more and more researches on the methods of hyperspectral image classification, GCN as a kind of network dealing with graph structure, the essence of which is to fully extract the long-range spatial dependencies between nodes, and the hyperspectral data can be transformed into a graph structure by using the nearest neighbor relationship between nodes in the image. Therefore, in this paper, the whole hyperspectral image is obtained as a global feature map by KNN method, and a feature map F_{KNN} of size 11×11×512 is output, and the whole feature map is transformed into a neighbor graph G. The GCN makes the information between each node pass in the graph, and then learns the neighbor information between the nodes in the graph to obtain the spatial topological relationship in the image. It is assumed that each local feature $\boldsymbol{x_i} \in R^{N\times 512}$ on the feature graph F_{KNN} is treated as a node on G, so all the local features will form a complete adjacency graph. any two nodes $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in G are connected by edges, and the correlation between all nodes is defined by the adjacency matrix $\boldsymbol{A} \in R^{N\times N}$. If the distance between two nodes is within a certain range, an edge is generated between these two nodes, and the smaller the distance, the stronger the correlation between the two nodes. In this paper, we first calculate the correlation $\boldsymbol{dis}\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right)$ between all nodes two by two to get the distance matrix $\boldsymbol{dis} \in R^{N\times N}$,

$$\boldsymbol{dis}\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) = \sqrt{\sum_{d=1}^{D}\left(\boldsymbol{x_{id}} - \boldsymbol{x_{jd}}\right)^2} \tag{3}$$

According to the KNN algorithm to sort Dis from smallest to largest, the K nodes closest to each node are selected as neighboring nodes, and edges are generated between these nodes, from which the adjacency matrix *A* is obtained. $\boldsymbol{A} \in R^{N\times N}$ is defined as.

$$\boldsymbol{A_{ij}} = \begin{cases} 1 & \boldsymbol{x_i} \in \boldsymbol{KNN}\left(\boldsymbol{x_j}\right) \text{ or } \boldsymbol{x_j} \in \boldsymbol{KNN}\left(\boldsymbol{x_i}\right) \\ 0 & \boldsymbol{other} \end{cases} \tag{4}$$

With the adjacency matrix *A* known, create the graph Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, where D is

the diagonal matrix representing the degree of A, i.e. $\boldsymbol{D}_{i,j} = \sum \boldsymbol{A}_{i,j}$; to enhance the generalization of the graph, the symmetric normalized Laplacian matrix L can be used, calculated as follows.

$$\boldsymbol{L} = \boldsymbol{I_n} - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}} \tag{5}$$

A spectral decomposition is performed on *L*, $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}$, where *U* is an orthogonal matrix consisting of eigenvectors and $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u_n})$ is the eigenset of *L*; $\boldsymbol{\Lambda}$ is the diagonal array of eigenvalues of *L*.

Defining the GCN as the product of the image signal $x \in R^N$ in the Fourier domain and the filter $g_\theta = diag(\theta)$, calculated as

$$x \cdot g_\theta = \boldsymbol{U} g_\theta \boldsymbol{U^T} x \tag{6}$$

where g_θ can be considered as a function of the eigenvalues ($\boldsymbol{\Lambda}$) of *L* with respect to the variables θ, i.e., $g_\theta(\boldsymbol{\Lambda})$, $\boldsymbol{U^T} x$ is the graph Fourier transform of *x*.

The Chebyshev polynomial T_k is used to approximate $g_\theta(\boldsymbol{\Lambda})$, denoted by $\lambda_{\max}$ as the maximum eigenvalue of L. Let $\hat{\boldsymbol{\Lambda}} = \frac{2}{\lambda_{max}}\boldsymbol{\Lambda} - \boldsymbol{I}_n$, i.e.

$$g_\theta(\boldsymbol{\Lambda}) \approx \sum_{k=0}^{K} \theta_k T_k\left(\hat{\boldsymbol{\Lambda}}\right) \tag{7}$$

$\hat{\boldsymbol{L}} = \frac{2}{\lambda_{\max}}\boldsymbol{L} - \boldsymbol{I}_n$ is the corresponding Laplacian matrix scaling, which is approximated by combining the constant equation $\left(U\boldsymbol{\Lambda}U^T\right)^k = U\boldsymbol{\Lambda}^k U^T$, as

$$x \cdot g_\theta = \sum_{k=0}^{K} \theta_k T_k\left(\hat{\boldsymbol{L}}\right) x \tag{8}$$

In this paper, we choose $K = 1$, i.e., only the 1st-order connectivity information of the nodes is considered, so that the normalized maximum eigenvalue $\lambda_{\max} = 2$, expanded and simplified as

$$\begin{aligned} x \cdot g_\theta &= \sum_{k=0}^{1} \theta_k T_k\left(\hat{\boldsymbol{L}}\right) x = \left(\theta_0 + \theta_1 \hat{\boldsymbol{L}}\right) x \\ &= \left[\theta_0 + \theta_1\left(\boldsymbol{L} - \boldsymbol{I}_n\right)\right] x \end{aligned} \tag{9}$$

where $\hat{\boldsymbol{L}} = \boldsymbol{L} - \boldsymbol{I}_n$, $\boldsymbol{L} = \boldsymbol{I}_n - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}$, and $\boldsymbol{I}_n$ are unit matrices, and to further simplify,

let $\theta_0 = -\theta_1 = \theta$, the above equation be equivalent to

$$x \cdot g_\theta = \theta \left(\tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tilde{\boldsymbol{A}} \tilde{\boldsymbol{D}}^{-\frac{1}{2}} \right) x \tag{10}$$

where, θ is the learnable weight; $\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_n$ is the normalized adjacency matrix and $\tilde{\boldsymbol{D}}_{ii} = \sum \tilde{\boldsymbol{A}}_{ij}$ is the normalized degree matrix.

A multi-layer GCN is eventually represented as

$$\boldsymbol{H}^{(l+1)} = \sigma \left(\tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tilde{\boldsymbol{A}} \tilde{\boldsymbol{D}}^{-\frac{1}{2}} \boldsymbol{H}^{(l)} \boldsymbol{W}^{(l)} \right) \tag{11}$$

Where, $\boldsymbol{H}^{(l)} \in \boldsymbol{R}^{\boldsymbol{N} \times d_l}$ denotes the node features of layer l; $\boldsymbol{H}^{(\boldsymbol{l}+1)} \in \boldsymbol{R}^{\boldsymbol{N} * d_{l+1}}$ denotes the updated node features. The features of the input layer are $\boldsymbol{H}^{(0)} = \boldsymbol{X}$. $\boldsymbol{W}^{(\boldsymbol{l})} \in \boldsymbol{R}^{d_l \times d_{l+1}}$ is the trainable weight in each layer, and $\sigma(\cdot)$ is denoted as the ReLU activation function.

The graph convolution operation is performed on *A* according to equation (11), and the features of each node are updated to obtain the contextual information with dimension 256 through the three-layer GCN.

2.3 Convergence and classification of 3D-DC and GCN

The features extracted by a single model are limited and cannot express the complete information of hyperspectral images, which eventually leads to unsatisfactory classification results, for example, 3D-DC ignores the topological relationship of space and GCN ignores the spectral information. In order to solve the above problems and make the feature extraction more comprehensive, this paper proposes to fuse 3D-DC and GCN to enhance the feature expression, and adopt the cascade fusion (Connnect Fusion) to fuse diverse features before the final classification, and the network structure is shown in Figure 2. Unlike the conventional network, in this paper, the empty spectral features of dimension 64 obtained from the 3D-DC network module and the contextual features of dimension 256 obtained from the three-layer GCN processing are pooled globally on average, and then the empty spectral features are cascaded with the spatial contextual features to obtain a feature of dimension 320. The fused features are processed by two layers of full concatenation before using Softmax classification to construct an end-to-end fusion network, and finally the whole network is trained using the cross-entropy loss function.

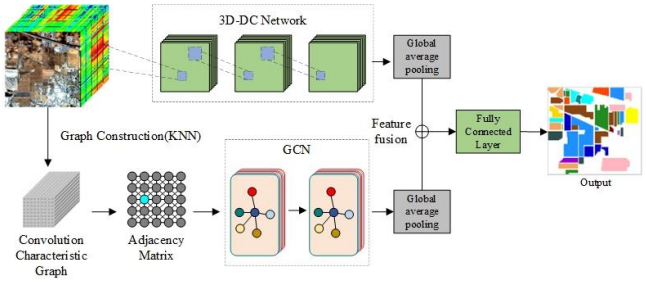

Fig. 2. Dilated-3D-GCN network structure

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Experimental Data Set

To evaluate the classification effectiveness of the proposed network structure, two benchmark datasets, Indian Pines (IP) and Pavia University (PU), are selected as the experimental data to validate the proposed method.

(1) The Indian Pines dataset was captured with the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor in 1992 at a remote sensing test site in northwest India with a spatial resolution of 20 m. There are 220 bands with a wavelength range of 0.4 to 2.5 nm, and the image size is 145 pixel × 145 pixel, containing a total of 16 categories of features. The pseudo-color images (bands 40, 80, and 120), ground truth maps, and each category name for this dataset are shown in Figure 3.

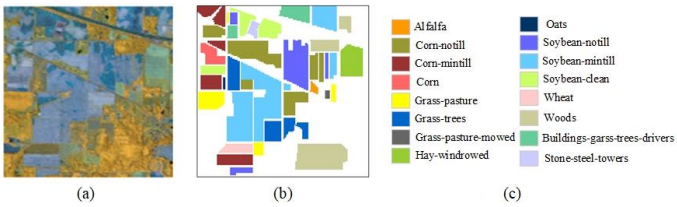

Fig. 3. Indian Pines dataset (a)Pseudo color image (b)ground truth map (c)classes

(2) The University of Pavia dataset was collected by the Reflection Optical System Imaging Spectrometer (ROSIS) sensor over the University of Pavia, Italy, in 2001, with a spatial resolution of 1.3 m, the presence of 115 bands, a wavelength range of 0.43 to 0.86 nm, and an image size of 610 pixel × 340 pixel, containing a total of 9 categories of features. The pseudo-color images (bands 10, 60, and 80), ground truth maps, and the names of each category for this dataset are shown in Figure 4.

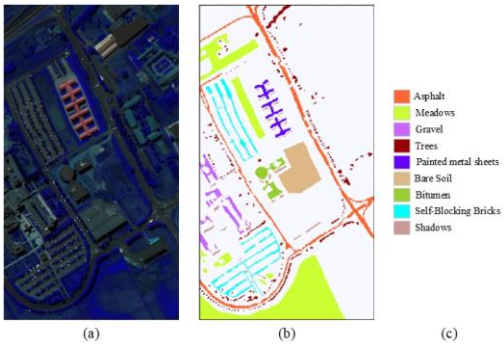

Fig. 4. University of Pavia dataset (a)Pseudo color image (b)ground truth map (c)classes

3.2 Evaluating Indicator

To verify the effectiveness of the proposed method, overall classification accuracy (OA), average accuracy (AA), and Kappa coefficient are used as the evaluation indexes of classification performance to evaluate the performance of hyperspectral image classification. Among them, OA indicates the ratio of the number of correct samples to the number of overall classification samples; AA is the mean value of the samples correctly predicted by each category after the model makes predictions; Kappa coefficient indicates the difference between the classification results of the model and the completely random classification results. Combining the above index factors, to ensure the randomness of the experimental data selection, 5%, 90%, and 5% data sample proportions were selected as the training, testing, and validation samples for each category of features in the IP data set, and similarly 1%, 98%, and 1% proportions were used as the training, testing, and validation samples for each category of features in the PU data set.

3.3 Parameter Analysis

Based on the determined network structure, the Adam optimizer is selected to train the network for optimal learning, setting the initial learning rate to 0.001 and the maximum number of iterations for network training to 200. Meanwhile, the main parameters are analyzed.

(1) The influence of void factor: The size of the dilated coefficient determines the range of the dilated convolution field, which can affect the extraction ability of the spatial features of hyperspectral images. For this reason, we set the dilated coefficients from 1 to 6 and analyze the effects of different values of dilated coefficients on the classification results in six groups of experiments, and the results are shown in Figure 5. The results are shown in Fig. 5. It can be seen from Fig. 5 that as the null coefficients increase, the range of perceptual fields becomes larger, and the features that can be learned by the convolutional layer become richer, thus improving the classification accuracy; however, when the null coefficients are large enough, some detailed information may be ignored or noise may be introduced, leading to a decrease in classification accuracy. For the IP and PU datasets, the classification accuracy is optimal when the hole coefficients are 2 and 4, with OA of 97.83% and 98.12%, respectively.

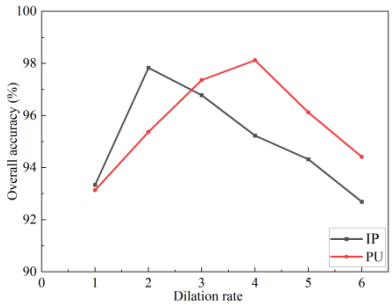

Fig. 5. Impact of expansion coefficient on classification accuracy in IP and PU datasets

(2) The Influence of the number of adjacent points K: The number of neighboring points K is the main parameter that constitutes the neighboring matrix, and the quality of the neighboring matrix determines the performance of the GCN, thus affecting the stability of the entire hyperspectral image spatial topology. Therefore, in this paper, the influence on the classification results is analyzed by selecting a total of four different numbers of neighboring points, 10, 15, 20, 25 and 30, and the results are shown in Figure 6. From Fig. 6, it can be seen that the classification accuracy is optimal when the number of neighboring points is 15 and 20 for IP and PU data sets, respectively, with OA of 99.14% and 99.26%, respectively. This is because when the number of neighboring points is small, the spatial contextual relationship of hyperspectral images cannot be expressed completely; however, when the number of neighboring points is too large, it is easy to cause redundancy or aggregation of node features that are not particularly relevant, resulting in a decrease in classification accuracy.

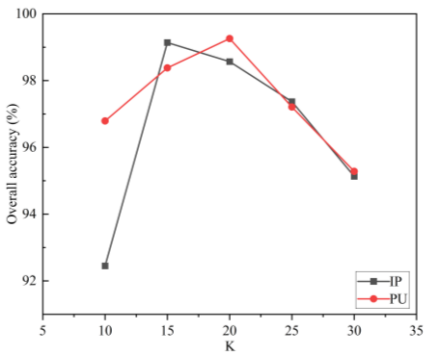

Fig. 6. Impact of the number of adjacent points K on the classification accuracy of IP and PU datasets

3.4 Ablation Experiment

In order to verify the effectiveness of end-to-end multi-scale feature fusion between the deep null-spectrum features extracted by the null convolution in 3D-DC networks and the contextual features extracted by the GCN module, ablation experiments are conducted on the IP and PU datasets in this paper. A three-dimensional convolutional neural network (3D-CNN) is used as the benchmark model in the experiments, and the network structure of all models is the same to ensure the accuracy of the experiments. The two comparison models and the Dilated-3D-GCN

model proposed in this paper are obtained by replacing some convolutional layers of the benchmark model with null convolution and adding the GCN module without KNN method composition to the benchmark model, respectively. The results of the overall precision classification versus training time for different network models are given in Table 1, respectively.

Table 1 Comparison of overall accuracy and training time of different network models

Methods	IP		PU	
	OA	***Training time /s***	***OA***	***Training time /s***
3D-CNN	93.24	214.6	94.23	178.4
3D-CNN-GCN (without KNN)	95.57	326.3	97.34	246.5
Dilated-3D-GCN (without KNN)	98.62	336.8	98.93	264.6
Dilated-3D-GCN	99.21	312.7	99.32	237.8

The following conclusions can be drawn from the analysis of the experimental results: (1) After adding the GCN module to the benchmark model, the training time increases while the classification accuracy improves by 2.34% and 3.11%, respectively. (2) Taking the experimental results of 3D-CNN-GCN (without KNN) and Dilated-3D-GCN (without KNN) as examples, the classification accuracy of the model improves by 3.05% and 1.29% on the IP and PU datasets, respectively, with almost the same training time after the introduction of the null convolution. (3) The classification accuracy of the proposed model (Dilated-3D-GCN) improves by 5.97% and 5.09% on the IP and PU data sets, respectively, by introducing both the null convolution and the GCN module in the presence of the KNN method in the benchmark model. (4) The training time of the model proposed in this paper is reduced by 13.6s, 24.1s, 8.7s and 26.8s compared to 3D-CNN-GCN (without KNN) and Dilated-3D-GCN (without KNN) on IP and PU datasets, respectively. The above experimental results illustrate that the proposed method using the null convolution and GCN module can improve the classification performance of the network model, proving the effectiveness of the multi-scale feature fusion of the null convolution and GCN module, while applying the KNN method to the graph convolution module to frame the graph can improve the classification efficiency of the model.

3.5 Results and Analysis

In order to prove the effectiveness and feasibility of this paper's method, this paper does experimental comparison with a total of five methods, namely, 3D-CNN, HybridSN, Dilated-3D-CNN, MDGCN, and CEGCN, respectively. Also, to ensure the fairness of the experiments, all five comparison algorithms are configured according to the optimal parameters in this paper, and the number of samples in the training set, validation set and test set are selected equally. Tables 2 and 3 show the classification accuracy of DilateDilated-3D-GCN and the comparison methods on the IP and PU data sets, respectively. The analysis and comparison in Table 2 and Table 3 show that (1) the best classification accuracy results are achieved by the method in this paper, in which the OA of Dilated-3D-GCN is improved by 6.15%, 2.55%, 1.49%, 0.82%, 0.44% compared with other methods on the IP dataset, and on the PU dataset, the OA of DilateDilated-3D-GCN improved by 14.89%, 2.11%, 1.19%, 0.58%, and 0.40%, respectively,

compared with other methods. (2) The limited features extracted by 3D-CNN network structures lead to low classification accuracy, but HybridSN improves classification accuracy by combining 2D-CNN with 3D-CNN and Dilated-3D-CNN by introducing null convolution. (3) The classification accuracies of MDGCN, CEGCN, and Dilated-3D-GCN network models are better than those of CNN-based network models, indicating that graph convolution can fully extract the contextual information of spatial topological relationships between irregular regions in images, and the accuracy is significantly improved on HSI classification. (4) Among the three GCN-based network models, the method in this paper takes advantage of 3D-DC and GCN to fuse the extracted deep spatial spectral features with the spatial contextual features in multiple features to achieve end-to-end network training, and the proposed method has the highest classification accuracy compared with the other three GCN-based methods.

Table 2 Comparison of overall accuracy and training time of different network models

Num	**Classification accuracy**					
	3D-CNN	***HybridSN***	***Dilated-3D-CNN***	***MDGCN***	***CEGCN***	***Dilated-3D-GCN***
1	81.25	89.23	76.63	92.35	**100.00**	**100.00**
2	88.73	95.73	96.34	98.12	98.25	**99.21**
3	87.98	93.27	**99.63**	98.23	99.22	98.57
4	88.67	95.74	99.34	99.28	98.24	**100.00**
5	97.63	97.25	**100.00**	**100.00**	98.64	98.18
6	93.96	98.64	99.42	**99.54**	99.84	98.06
7	87.23	93.19	96.83	97.73	**100.00**	**100.00**
8	**100.00**	99.56	**100.00**	**100.00**	**100.00**	**100.00**
9	75.56	**100.00**	85.83	**100.00**	**100.00**	82.32
10	88.22	96.52	95.63	98.83	**99.33**	98.21
11	93.42	96.21	95.62	98.14	98.83	**99.21**
12	94.89	93.13	**100.00**	98.58	96.69	99.42
13	**100.00**	**100.00**	97.67	**100.00**	99.64	**100.00**
14	97.86	98.93	**100.00**	99.84	**100.00**	99.23
15	84.37	94.31	97.82	96.21	98.86	**100.00**
16	**100.00**	**100.00**	92.26	97.12	91.21	**100.00**
OA/%	93.12	96.72	97.78	98.45	98.83	**99.27**
AA/%	92.52	95.99	96.25	97.82	98.48	**98.81**
Kappa/%	92.94	95.48	96.79	97.97	98.65	**99.12**

Table 3 Comparison of overall accuracy and training time of different network models

Num	Classification accuracy					
	3D-CNN	***HybridSN***	***Dilated-3D-CNN***	***MDGCN***	***CEGCN***	***Dilated-3D-GCN***
1	90.23	94.74	97.84	96.53	99.24	**99.47**
2	98.32	99.45	99.14	**99.46**	99.35	99.92
3	87.31	98.45	93.72	98.34	98.75	**98.75**
4	92.31	88.58	**99.32**	96.15	98.23	99.12
5	85.68	99.91	**100.00**	99.45	97.96	96.27
6	95.38	99.78	98.36	**100.00**	97.53	**100.00**
7	88.34	99.74	95.67	99.49	**100.00**	99.24
8	88.11	92.67	92.37	96.23	97.67	**98.57**
9	99.24	89.33	99.28	97.53	98.94	**100.00**
OA/%	94.43	97.21	98.13	98.74	98.92	**99.32**
AA/%	92.65	96.45	97.31	97.88	98.45	**99.12**
Kappa/%	92.15	96.21	97.75	98.23	97.85	**98.38**

Table 4 shows the training time and testing time of different methods on both IP and PU datasets. From Table 4, the GCN-based deep learning methods increase the time to construct the graph adjacency matrix to some extent, the training times are both longer than the CNN-based deep learning methods. Among the GCN-based methods, the method in this paper has the shortest time cost. This is because the method uses KNN to extract features for the construction of the adjacency graph, which effectively reduces the composition time. Comprehensive analysis of the comparison results of training time and classification accuracy can conclude that the network model in this paper can achieve the best results.

Table 4 Training time of different network models

Dataset	***Time***	***3D-CNN***	***HybridSN***	***Dilated-3D-CNN***	***MDGCN***	***CEGCN***	***Dilated-3D-GCN***
IP	Train/s	73.7	231.2	279.6	356.7	389.3	468.7
PU	Test/s	23.2	98.3	143.7	181.3	235.8	85.2

4 CONCLUSION

To address the problems that CNN has large network parameters and can only extract features for regular regions of fixed size, which leads to insufficient information extraction; meanwhile, GCN tends to extract features in the same way for neighboring nodes and ignores the changing regions existing in hyperspectral images, etc., this paper proposes a hyperspectral image classification method based on three-dimensional dilated convolution and graph convolution (Dilated-3D- GCN). This method firstly, the image is extracted with deep null-spectral features by 3D-DC module; Secondly, extract the spatial context of the image using KNN to extract the global information of the image as input to the convolution; finally, the global features extracted by 3D-DC module are fused with the spatial contextual features extracted by GCN module to improve the utilization of information in the image and balance the problems of CNN and GCN in hyperspectral images. In this paper, experiments are conducted on two datasets, and the classification performance of this method and other advanced methods are evaluated and analyzed comprehensively using three evaluation metrics. The experiments

finally show that the proposed Dilated-3D-GCN method achieves 99.27% and 99.32% on the IP and PU datasets, respectively. This shows that the method not only extracts the deep null-spectral features more completely, but also extracts the spatial topological relations by GCN using the null-spectral features, and the fusion of multiple features effectively improves the classification results and the information expression ability, which shows certain superiority in the actual feature classification.

REFERENCES

[1] Wei W, Zhang L, Li Y, et al. Intraclass similarity structure representation-based hyperspectral imagery classification with few samples[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 1045-1054.

[2] Ablet E, Maimaitiaili B, Sawut M, et al. Combined estimation of chlorophyll content in cotton canopy based on hyperspectral parameters and vack Propagation neural network[J]. Acta Optica Sinica, 2019, 39(9): 0930003.

[3] Li Y S, Li H K, Xu F. Research progress of hyperspectral remote sensing monitoring in mine environmental contamination[J]. Nonferrous Metals Science and Engineering, 2022, 13(1): 108-114.

[4] Qian S E. Hyperspectral satellites, evolution, and development history[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 7032-7056.

[5] YE Z, BAI L, HE M Y. Review of spatial-spectral feature extraction for hyperspectral image [j]. Journal of Image and Graphics, 2021, 26(8): 1737-1763.

[6] WANG J Y, ZHANG C M, ZHANG Y B, et al. Semi-supervised classification algorithm of hyperspectral image based on DL1 graph and KNN superposition graph. Sci Sin Inform, 2017, 47: 1662-1673.

[7] WANG Y, YU W, FANG Z. Multiple kernel-based SVM classification of hyperspectral images by combining spectral, spatial, and semantic information[J]. Remote Sensing, 2020, 12(1): 120-145.

[8] WU F Y, WANG X,DING J W,et al. 2020. Improved cascade forest deep learning model for hyperspectral imagery classification. Journal of Remote Sensing, 24(4): 439-453.

[9] YU A Z, LIU B, XING Z P, et al. Salient feature extraction method for hyperspectral image classification[J]. Acta Geodaetica et Cartographica Sinica, 2019, 48(8): 985-995.

[10] HUANG H, WANG L H, SHI G Y. Spatially-Regularized Manifold Discriminant Analysis Algorithm for Hyperspectral Image Classification[J]. Acta Optica Sinica, 2020, 40(2): 0228001.

[11] BEIRAMI B A, MOKHTARZADE M. Band grouping SuperPCA for feature extraction and extended morphological profile production from hyperspectral images[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(11): 1953-1957.

[12] PAN C, JIA X, LI J, et al. Adaptive edge preserving maps in Markov random fields for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(10): 8568-8583.

[13] JIA S, ZHUANG J, DENG L, et al. 3-D Gaussian-Gabor feature extraction and selection for hyperspectral imagery classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11): 8813-8826.

[14] WAN Y L, ZHONG X W, LIU H, et al. Survey of application of convolutional neural network in classification of hyperspectral images[J]. Computer Engineering and Applications, 2021, 57(4): 1-10.

[15] LIU X, SUN Q, MENG Y, et al. Feature extraction and classification of hyperspectral image

based on 3D-convolution neural network[C]//2018 IEEE 7th data driven control and learning systems conference (DDCLS). IEEE, 2018: 918-922.
[16] ROY S K, KRISHNA G, DUBEY S R, et al. HybridSN: Exploring 3D-2D CNN feature hierarchy for hyperspectral image classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(2): 277-281.
[17] JEYARAJ P R, NADA E R S, PANIGRAHI B K. ResNet convolution neural network based hyperspectral imagery classification for accurate cancerous region detection[C]//2019 IEEE Conference on Information and Communication Technology. IEEE, 2019: 1-6.
[18] CHEN W H, HE J, LIU G. Hyperspectral image classification based on convolution neural network with attention mechanism. Laser & Optoelectronics Progress: 1-15[2022-10-28]. http://kns.cnki.net/kcms/detail/31.1690.TN.20210802.1730.070.html
[19] YAN M J, SU X Y. Hyperspectral image classification based on three-dimensional dilated convolutional residual neural network[J]. Acta Optica Sinica, 2020, 40(16): 1628002.
[20] HEIDARI N, IOSIFIDIS A. Progressive graph convolutional networks for semi-supervised node classification[J]. IEEE Access, 2021, 9: 81957-81968.
[21] WAN S, GONG C, ZHONG P, et al. Multiscale dynamic graph convolutional network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(5): 3162-3177.
[22] ZUO X B, LIU B, YU X C, et al. Graph convolutional network method for small sample classification of hyperspectral images[J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(10).
[23] LIU Q, XIAO L, YANG J, et al. CNN-enhanced graph convolutional network with pixel-and superpixel-level feature fusion for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(10): 8657-8671.